

Written assignment to the Medical Faculty, University of Oslo:

Title:

The Norwegian version of Mini-International Neuropsychiatric Interview (M.I.N.I.):
Feasibility, patient acceptability and test-retest reliability in an acute psychiatric ward.

Running head:

M.I.N.I.: feasibility, patient acceptability and test-retest reliability.

Corresponding author:

Øystein Gundersen

Medical student, H-02

University of Oslo

Boks 1072 Blindern

0316 OSLO

Norway

Phone: +47 936363608

E-mail: oeysteingundersen@hotmail.com

Instructors:

Cand.med., researcher, Jon Mordal, Lovisenberg diaconal Hospital

Dr.med., researcher, Jørgen G. Bramness, Norwegian Institute of Public Health

ABSTRACT

Background:

Mini-International Neuropsychiatric Interview (MINI) is a structured psychiatric diagnostic interview. Feasibility, patient acceptability, reliability and validity of MINI have been tested in other countries, but not yet in Norway.

Objective:

The aim of the present study was to test the feasibility, patient acceptability and test-retest reliability of the Norwegian MINI version in an acute psychiatric ward.

Methods:

From August 2006 to February 2007 38 patients were interviewed with MINI in the acute psychiatric ward at Lovisenberg Diaconal Hospital. Each of the patients was interviewed twice by two different interviewers, with a mean delay of 1.8 days. All interviews were timed to assess feasibility, and the patients filled out a response form to assess the patient acceptability. Cohen's kappa was calculated to investigate the test-retest reliability.

Results:

Mean interview time was 34 minutes indicating satisfactory feasibility. However, this is a longer duration than reported in other studies. The results also indicate that MINI was accepted by the patients. Test-retest reliability was quite high, indicating good or very good agreement for 11 of 16 MINI disorders. However, 2 of the MINI diagnoses had low kappa.

Conclusion:

The overall results were positive. Difficulties in interviewing patients with severe psychopathology may be the cause of the few low kappa values and the longer mean duration of the interview.

1 INTRODUCTION

1.1 Psychiatric diagnoses

Psychiatric diagnoses are in clinical practice usually set on basis of a clinical interview. Psychiatry as a medical discipline separates from most other medical disciplines in the lack of objective measures in the diagnostic process. Diagnoses in other disciplines may also be based on clinical interviews, but this information usually supplements the clinical tests and laboratory or radiology results (Pinninti et al. 2003). Although the progress in biological psychiatry may provide tools to support diagnoses, psychiatric diagnoses can not at present be calculated exactly. The human mind with emotions, thoughts and believes are abstract values that may not ever be exactly measured (Malt, Retterstøl, & Dahl 2003b). Exceptions from this may be the organic psychiatric disorders, for example organic psychosis, where a somatic cause of the psychiatric symptoms that are presented may be found (Malt, Retterstøl, & Dahl 2003c).

The International Statistical Classification of Disease and Related Health problems, ICD-10, and The Diagnostic and Statistical Manual of Mental Disorders, DSM-IV, list different categories of mental disorder and the criteria for these diagnoses. ICD-10 is developed by WHO and chapter F is about psychiatric and behavioural disorders (1992, Norwegian edition 1996). DSM-IV (1994, 1996) is composed by the American Psychiatry Association and is not an official tool in Norway. Although WHO and the American Psychiatry Association have attempted to harmonize the two classification systems, there are still some differences (Malt, Retterstøl, & Dahl 2003c). ICD-10 is the most frequently used system across the world for clinical work and training purposes, while DSM-IV is the most frequently used for research work (Zimmerman et al. 2005). ICD-10 is used in Norway since 1999 for diagnoses of psychiatric disorders.

1.2 Psychometric tests

In unstructured psychiatric interviews clinicians will obtain different kinds of information, and the diagnoses and its severity may be evaluated differently. This may inherit potential sources of error. Use of psychometric tests may reduce this problem (Malt, Retterstøl, & Dahl 2003a). Psychometric tests may be divided by form (interviews versus self-reports), and by the content (diagnose versus disorder severity) (Blacker 2005).

1.2.1 Diagnostic structured interviews

For clinical and research purposes, extensive interviews have been developed. Two of the most used structured interviews in this category are SCID (Structured Clinical Interview for DSM-IV) (First et al 1997) and SCAN (Structured Clinical Interview for Neuropsychiatry) (WHO 1992) for ICD-10. Both SCID and SCAN are time consuming and the interviewer should be either a psychiatrist or a clinical psychologist with special training in the interview-technique. CIDI (Composite International Diagnostic Interview) also belongs to this category and is more structured than SCID and SCAN. It is based on yes and no answers and therefore more suited for non-academic settings. CIDI, SCAN and SCID contain more than 500 questions, and each of these requires at least 1-2 h to conduct (Malt, Retterstøl, & Dahl 2003a;Sheehan et al. 1997).

Diagnostic interviews for the primary health care includes the Norwegian SPIFA (Structured interview for the primary health care) (Dahl et al 2003) and Prime-MD (Primary Care Mental Disorders) (Spitzer et al 1994). The doctor only needs a short introduction to the interview. SPIFA covers 22 psychiatric disorders. In Prime-MD the patient first answers 25 questions, and the doctor uses these answers to ask new and more specific questions, for screening the six most common psychiatric disorders in primary care (Malt, Retterstøl, & Dahl 2003a).

1.2.2 MINI

The MINI International Neuropsychiatric Interview was developed by Sheehan and Lecrubier in 1990. MINI was developed to meet the need for a brief, reliable and valid structured diagnostic interview that screens many disorders. MINI is based on yes and no answers, contains 16 sections and screens axis 1 DSM-IV disorders and also anti-social personality disorder; all together 27 past and current disorders. It is organized in diagnostic sections and uses branching tree logic; it has two to four screening questions per disorder. Additional symptom questions are asked only if the screen questions are positively endorsed. The composition of MINI makes it easier to be used also by non-academics. This fact, along with its time saving aspect, makes it ideal in large-scale, multi-centre clinical trials for example in psychopharmacology, or for clinical work in a psychiatric ward (Lecrubier et al. 1997;Sheehan et al. 1998a).

For the English version of MINI, excellent inter-rater and test-retest reliability, and moderate validity of MINI versus CIDI (Lecrubier et al. 1997) and SCID (Sheehan et al. 1997) have been reported. MINI has been translated into 43 different languages (cited Sep.

20, 2007; available from: www.medical-outcomes.com/HTMLFiles/MINI/MINI.htm) and validity and reliability has been explored for the French (Lecrubier et al. 1997), Japanese (Otsubo et al. 2005), Italian (Barbui C. et al. 2004) and Moroccan (Kadri et al. 2005) versions. MINI is now the most used structured psychiatric diagnostic interview in the world (cited Sep. 20, 2007; available from: www.medical-outcomes.com/HTMLFiles/MINI/MINI.htm).

The “MINI family” consists of different versions of MINI. MINI-Plus is the most detailed version. The format of this interview is still less complex than that of SCID, SCAN and CIDI. MINI-Screen is a screening instrument for primary care covering more than the six disorders tapped by the PRIME-MD, but retaining the brevity of these instruments. MINI-Kid is a structured instrument for child and adolescent psychiatry (Sheehan et al. 1998b).

A Norwegian version of the MINI was translated by Kari Ann Leiknes et al in 1999. The Norwegian 5.0.0 version was published in March 2006. In 2002 The Norwegian Psychiatry Association formed a panel for acute psychiatry. This panel was supposed to work out common guidelines for the diagnostic process in acute psychiatry. These guidelines say that MINI should be part of the standard procedure in all first time admissions (from “utredning av akuttpsykiatriske tilstander” produced by Utvalg for akuttpsykiatri). MINI is used in many different settings in Norway. However, the validity and reliability of the Norwegian version are not yet investigated. Considering the guidelines, this would be of interest.

1.2.3 Interviews that score the severity of a psychiatric disorder

Examples of this are Montgomery Åsberg Depression scale (MADRS), that scores severity of depression, and Positive and Negative Syndrome Scale (PANSS) that scores positive and negative symptoms in schizophrenia/psychosis (Blacker 2005; Malt, Retterstøl, & Dahl 2003d).

1.3 Testing diagnostic structured interviews

Structured interviews must combine four characteristics to be useful in clinical practice: They should be feasible, acceptable to patients and have a high validity and reliability (Pinninti et al. 2003).

1.4.1 Feasibility

Feasibility can be defined as “capability of being done; practicability” (cited Sep. 19, 2007; available from: www.oed.com). Structured interview feasibility most commonly relates to interview length (Pinninti et al. 2003). Acceptable duration of a structured interview is related

to how long psychiatric patients can concentrate, and the time available to be used for this purpose in a clinical reality. MINI is usually performed in approximately 15 minutes (cited Sep. 20, 2007; available from: www.medical-outcomes.com/HTMLFiles/MINI/MINI.htm). This is of course dependent on both the capability of the interviewers and the patients' degree of psychopathology.

1.4.2 Patient acceptability

“Patient acceptance relates to the avoidance of ambiguous or complex phrasing of questions, the gradual rather than abrupt transitioning from one symptom area to another, and the type of response format utilized” (Pinninti et al. 2003). Hence, the patient acceptance is dependent of what kind of questions that are asked in the response format and how the patients are supposed to answer. On questions about how satisfying a patient experienced a structured interview, patients can for example be given two options, yes (satisfying) and no (not satisfying), or be given the option to range the satisfaction on a scale from 0 to 10 or making a cross on a line from minimum to maximum satisfaction (Visual analogue scale; VAS-scale) (cited Sep. 21, 2007; available from: [painworld.zip.com.au/articles/tools/Visual%20Analogue%20Scale%20\(VAS\).pdf](http://painworld.zip.com.au/articles/tools/Visual%20Analogue%20Scale%20(VAS).pdf)).

1.4.3 Validity

Validity can be defined as “the extent to which a measurement, test or study measures what it purports to measure” (Anderson et al. 1994). When we evaluate if a measurement is valid, we need a “gold standard”. In the American article investing the validity of MINI, SCID was used as the gold standard (Sheehan et al. 1997). MINI's diagnoses were also tested against the diagnoses made by an expert. Diagnostic concordance was assessed using unweighted kappa values, sensitivity, specificity, positive predictive value, and negative predictive value.

1.4.4 Reliability

Reliability can be defined as: 1. “The extent to which a statistically derived measure from a sample gives the same results upon repeated sampling under identical conditions”. 2. “The tendency of a system to be resistant to failure” (Anderson et al. 1994). Thus, the reliability of an instrument reveals whether the instrument is robust or not. Traditionally there are three different types of reliability (Friis & Vaglum 1999a) :

- Internal consistency reliability

This type of reliability can be used when you have a test where different scores are put together in a sum, for example in a test to determine the degree of anxiety or in a test to determine the degree of depression as in MADRS.

- Inter-rater reliability

This is a measurement of the agreement between two persons that evaluate the same patient.

- Test-retest reliability

This is a measurement of degree of consistency between scorings; when the same test is used twice with a gap of time between.

In testing inter-rater and test-retest reliability the coefficient Cohen's kappa is much used. The psychiatric diagnoses MINI screens for are measurements on nominal scales (Friis & Vaglum 1999b). A nominal scale has no stages and no zero. It is composed of different categories. Testing the reliability of measurements on nominal scales, it was earlier considered sufficient to calculate the degree of agreement using percentage. This way of calculating did however not take the chance agreement into account, and thus kappa was developed. This is also a measurement of agreement, but corrugated for chance agreement. Kappa has also got some weaknesses; categories that occur very often or very seldom in a material, will produce a low reliability value. The reliability can thus be under valued by the scholar, and it is therefore recommended to calculate two indexes measuring degree of agreement for presence and degree of agreement for absence of a phenomena (Friis & Vaglum 1999a). Inter-rater and test-retest reliability are investigated in for example the American MINI validity/reliability article using Cohen's kappa (Sheehan et al. 1997).

If diagnoses occur less than 5 % in the material, kappa is unreliable (Altman D.G 1991; Malt, Retterstøl, & Dahl 2003c).

Cohen has suggested cut-off values for interpretation of kappa (table 1). In practice, any value much below 0.5 indicates poor agreement. The degree of acceptable agreement must depend upon circumstances (Altman D.G 1991).

1.5 Psychiatric structured interviews in different countries and settings

1.5.1 Testing psychiatric structured interviews in different countries

It is of importance to test psychiatric structured interviews that are translated into different languages. Apparently identical questions in for example Norwegian and English will have a potential of meaning slightly different things. Questions in English can produce high validity and reliability while the same questions in Norwegian do not necessarily produce the same

validity and reliability. This does not just concern the semantic divergences, but also divergences in cultural understanding of words and sentences. A psychiatric symptom can be described in different ways in Norwegian, Moroccan and English and makes it relevant to test the cross-cultural validity and reliability. It is important to find expressions that give meaning to a psychiatric patient (cited Sep. 25, 2007; available from: en.wikipedia.org/wiki/Translation).

1.5.1 Testing psychiatric structured interviews in different settings

Structured psychiatric interviews can be used in different settings: In primary care, acute wards, intermediary wards and outpatient clinics. Ideally psychiatric structured interviews should be tested in all types of settings in all countries.

1.6 Aims of the study

We wanted to investigate the feasibility, patient acceptability and test-retest reliability of the Norwegian MINI version in an acute psychiatric ward.

2 MATERIAL AND METHODS

2.1 Population

The study was performed at Lovisenberg diaconal Hospital, acute psychiatric ward. The hospital is a major public hospital in Oslo, the capital of Norway. It provides psychiatric and medical services to 97 000 inhabitants of the inner city.

Included in this study were 38 patients (22 women), and the mean age was 38 years (range 22-71 years). They were interviewed twice. Two of the subjects did not complete one of the two interviews. ØG and JM was interviewer number one in 18 and 20 interviews respectively. Mean delay between interview number one and two was 1 day and 19 hours (Range 0-6 days).

2.2 Procedures

In the first part of the study, August 2006, we randomised patient inclusion. We also randomised which of the two interviewers, ØG or JM, should interview first. Patients that did not agree to be interviewed or of other reasons did not complete interviews were registered. In the second part, from September 2006 to February 2007, this study was part of a bigger project. In this period we did not randomise patient inclusion and did not register patients that did not want to participate.

2.3 Instruments and interviewers

To assess feasibility each interview was timed.

Patient acceptability was recorded by using a visual analogue scale (VAS-scale) ranging from 0 to 10 (Appendix 1); Zero indicating minimum acceptability and 10 indicating maximum acceptability. This is an acknowledged method implying use of a line with for example minimum satisfaction in one end of the line and maximum satisfaction in the other end of the line. The patient then set a mark on the line to express how satisfied they were with the procedure (cited Sep. 21, 2007; available from: [painworld.zip.com.au/articles/tools/Visual%20Analogue%20Scale%20\(VAS\).pdf](http://painworld.zip.com.au/articles/tools/Visual%20Analogue%20Scale%20(VAS).pdf)). This way of scoring the acceptance may give a more nuanced picture than just presenting two answer alternatives (yes/no). The questions were quite similar to the ones used in the article investing the acceptability of the English MINI version (Pinninti et al. 2003). However, they only used two answer alternatives. We made adjustments of these questions ourselves and used the VAS-scale.

Patient Question number 1 was: Did the questions cover all your problems? (0 = very poorly/totally useless, 10 = very good/very useful)

Patient Question number 2 was: In general; how did you experience this interview? (0 = very unpleasant, 10 = meaningful/pleasant)

Patient Question number 3 was: Do you have any comments? (Open ended)

To assess test-retest reliability, we used the Norwegian version of MINI 5.0.0 (appendix 2). Three of the sections are optional (major depressive disorder with melancholic features, posttraumatic stress disorder and antisocial personality disorder), and we chose not to include these to make the interviews as short as possible. We included the suicidality section in the interviews, but this section was not assessed in the analysis. In total, 24 lifetime/past/recurrent and current disorders were assessed in each interview and 23 were assessed in the analysis (table 2 and 3). All the interviews were performed by ØG and JM independently, with some time delay. The interviewers were blinded for each others results, and did not know the clinical diagnoses of the patients.

2.4 Statistics

We used SPSS version 14.0 to analyse the data material in our study. When assessing the relationship between two categorical variables, chi-square was used. When assessing the relationship between one categorical and one continuous variable, student's t-test was used. When assessing the relationship between two continuous variables, Pearson's r correlation coefficient was used. Cohen's Kappa was used to assess the test-retest reliability.

2.5 Ethics

The study was based on written informed consent (appendix 3). It will always be an issue in studies with psychiatric patients that they may not be in shape to protect their own integrity. This was in our minds throughout the project, and we tried to be sensitive to possible problems connected to this. The study was approved by the local ethical committee (REK sør) as part of a larger project, which in turn also was approved by the Norwegian Data Inspectorate. MINI is, both for clinical and research purposes, free for use.

3 RESULTS

3.1 Population and prevalence of disorders

The prevalence of all disorders is shown in table 2. Two of the second MINI interviews were not fully completed, but were included in the analyses. This was due to the fact that these two interviews were nearly finished and generated approximately the same number of diagnoses as the corresponding first interviews (0 and 2 in the first versus 1 and 2 in the second interviews). When 38 patients were assessed, the mean number of diagnoses in the first interview was 5.0 (3.8-6.1 95%CI, range 0-12), and in the second interview the mean number was 4.8 (3.8-5.9 95%CI, range 0-11). This difference was not significant (Students t-test, $p=0.836$). A difference of 3 or more diagnoses between interview 1 and 2 were observed in 5 patients. For the prevalence of specific diagnoses, there was no significant difference in interview 1 and 2.

In interviews performed by ØG, the mean number of diagnoses was 4.8 (3.7-5.9 95%CI, range 0-12) and by JM the mean number was 5.0 (3.9-6.0 95% CI, range 0-11). This difference was not significant (Students t-test, $p=0.836$). JM used two diagnostic categories significantly more often than ØG. This was for past hypomanic episode (6 versus 0, chi square $p=0.011$) and GAD (12 versus 3, $p=0.009$). For all other diagnoses, no significant differences were observed.

Of 76 interviews, the most prevalent diagnoses were major depressive disorder, current (67%), panic disorder, lifetime (49 %), psychotic disorder, lifetime (45%), major depressive disorder, recurrent (43%), agoraphobia (40%) and psychotic disorder, current (36 %). Only two patients did not fulfil criteria for any diagnose in both interviews.

3.2 Feasibility

The mean time to perform an interview was 34 minutes (31-36 min. 95%CI, range 12-69 min.). JM used on average 31 minutes (28-35 min. 95%CI, range 12-69 min.) while ØG used 36 minutes (32-40 min. 95%CI, range 15-60 min.). This difference was non significant (Student's t-test, $p = 0,111$) Mean time to perform the first interview, regardless of who performed it, was 36 minutes (32-40 min. 95%CI, range 12-69 min.) and the second was 31 minutes (28-35 min. 95%CI, range 15-56 min.). This difference was non significant (Pearson's $r = 0.499$). JM used on average 35 minutes to perform the first interview, while ØG used 37 minutes. This difference was also non significant (Student's t-test, $p = 0.684$). JM used on average 27 minutes on the second interviews, while ØG used 35. This difference was significant (Student's t-test, $p = 0.036$).

The time used for performing the MINI interview was not related to the interview number with only a slight negative correlation coefficient (Pearson's $r = -0.082$, NS). For one of the interviewers (JM) this trend was more pronounced, but still not significant (Pearson's $r = -0.157$, NS). See also figure 1.

The time for performing the MINI interview was borderline positively significant related to the number of diagnosis set (Pearson's $r = 0.230$, $p = 0.051$). This was due to a rather large relationship between the number of diagnosis set on the first interview and the time to perform the interview ($r = 0.339$, $p = 0.038$), while no such relationship was found on the next interview ($r = 0.096$, $p = 0.584$). See also figure 2.

3.3 Patient acceptability:

Patient acceptability forms were completed by 36 patients. On question 1, the mean result was 6.6 (5.6-7.7 95%CI, range 0.2-10) and on question 2, the mean was 7.9 (7.2-8.5 95%CI, range 1.5-10). The mean result on question 2 was not significantly correlated to age, sex, patientID, time of MINI, diagnose, number of diagnoses or name of interviewer (Student's t-test). On question 1, however, the 13 patients with current psychosis scored a higher mean value than the others (8.1 vs. 5.8, Student's t-test, $p = 0.036$). Also on question 1, patients with low total number of diagnoses scored lower mean value than others: The 5 patients with 0 diagnoses scored 3.7 vs. 7.1 (Student's t-test, $p = 0.021$), and the 15 patients with less than 5 diagnoses scored 5.3 vs. 7.6 (Student's t-test, $p = 0.027$). Finally, the 18 patients interviewed by JM had a higher mean value on question 1 than ØG (7.7 versus 5.5, Student's t-test, $p = 0.033$).

Additional comments were given by 24 patients. See table 4 for examples.

3.4 Test-retest-reliability:

Kappa values were calculated for 16 of the 23 MINI diagnoses (see table 3). They indicated very good agreement for 5 diagnoses (i.e. major depressive disorder, current; panic disorder, current; panic disorder, lifetime; drug dependence; alcohol dependence), good agreement for 6 diagnoses (i.e. major depressive disorder, recurrent; manic episode, past; agoraphobia; psychotic disorder; current; psychotic disorder, lifetime; bulimia), moderate agreement for 2 diagnoses (i.e. manic episode, current; social anxiety disorder), fair agreement for 2 diagnoses (i.e. generalized anxiety disorder (GAD) and dysthymia) and poor agreement for 1 diagnose (i.e. obsessive-compulsive disorder (OCD)). Due to prevalence below 5 %, the kappa value for bulimia nervosa should be assessed with care. Drug abuse, anorexia nervosa and anorexia nervosa, binge eating/purging type were excluded from analysis because none of the patients

met MINI criteria for these 3 diagnoses. We did not have sufficient sample size to analyze current and past hypomanic episode, alcohol abuse and mood disorder with psychotic features, and these 4 diagnoses were also excluded from analysis (none of the patients met MINI criteria in both interviews, producing negative kappa values).

4 DISCUSSION

This study indicated positive results both for patient acceptability, feasibility and test-retest reliability of the Norwegian version of MINI in an acute psychiatric ward. For most diagnoses, the kappa values indicated good or very good agreement. However, for some of the diagnoses the kappa values were lower, especially for OCD and GAD. The mean time of the interview was satisfactory, although manifesting some longer duration than the guidelines indicate.

Feasibility of the MINI has traditionally been estimated by timing the interview. The interview has been tested in many populations and usually takes a little below 20 minutes in average (Sheehan et al. 1997), but this is of course dependent on both interviewer and patient characteristics. Specifically interviewer experience with MINI may reduce the time to perform the interview. This may be supported by the fact that JM, however not ØG, had a slight reduction in time used from the first to the last interview in the project. ØG's lack of reduction may be due to less motivation to be quick because he was a student with more time available. Moreover, ØG's mean interview time was longer than JM's. This may indicate that clinical experience also makes the interview time shorter. The interviewers' focus – especially ØG's – was perhaps a bit one-sided on making the interviews as precise as possible in order to increase reliability. This may also have played a part in increasing time to perform interviews.

Patient characteristics may indeed influence the time it takes to administer the MINI interview. Severity of psychopathology may influence and may by itself explain this fact. The number of diagnosis may also influence the time used (Lecrubier et al. 1997). The mean number of diagnosis detected in our sample was 5 in the first interview and 4.8 in the second. These are relatively high numbers. There was a large positive relationship between the number of diagnosis detected in the first interview and the time to take the interview. In 36 % of the interviews the patients met the MINI criteria for current psychosis. This also supports that we are dealing with a group of patients displaying severe psychopathology. There are differences, concerning socioeconomic status, somatic and psychiatric health, between people living in different parts of the city (Thorsnæs et al. 2006b). Lovisenberg covers the inner part of Oslo and this population may display high degree of psychopathology (Thorsnæs et al. 2006a).

Still, the mean time of the interviews was much shorter than what could be expected for SCID, SCAN or CIDI (Sheehan et al. 1997). Thus, the time it takes to interview patients with more severe psychopathology does not invalidate the MINI as a well suitable instrument in these settings.

As far as we know, only one study has been conducted on the patient acceptability of MINI (Pinninti et al. 2003). Similar to our results, they found that MINI was accepted by the patients. However, this study used only yes and no alternatives. Ninety-four percent reported that MINI covered all their symptoms. They asked if the patients were bothered in any way by the questions, and 89 percent answered no. Our version of this question is more open (In general, how did you experience the interview), and in my opinion thus gives more information. In the present study, patients with psychosis current/earlier, scored higher on question 1. This may imply that the psychosis section in MINI covers essential psychotic symptoms, but is irrelevant and perhaps alienating to non-psychotic patients. The mean value on question 2 is very high. This may be due to a structured interview's ability to organize the chaos characterizing patients in acute psychiatric wards. Perhaps screening of many relevant symptoms will give patients an overview over themselves, help building or rebuilding a "self", and deliver words to express how they feel. Sometimes it is easier to answer yes and no than to construct sentences. This gives MINI a purpose in acute psychiatry independently of high or low reliability. Some of the patient comments may support this (table 4).

Sixty-nine percent (11/16) of our kappa values were good or very good. This is superior to the results in the Italian and Japanese articles. The kappa values in the Italian study in general were much lower than ours (Barbui C. et al. 2004). The Japanese study reported 67 percent (8/12) good or very good kappa values (Otsubo et al. 2005). The American article reported 91 percent (21/23) good or very good values, but due to changes in MINI there are differences from our study in diagnostic sections (Sheehan et al. 1997). In the French study only 5 diagnoses were assessed to improve the material, and all of the values were good or very good (Lecrubier et al. 1997). Usually, one person performs both interviews when investigating the test-retest reliability. The fact that we were two different interviewers in the first and second interview may indicate that our results are conservative. However, the other MINI articles, except the Moroccan, also have different interviewers in the first and second interview (Barbui C. et al. 2004;Kadri et al. 2005;Lecrubier et al. 1997;Otsubo et al. 2005;Sheehan et al. 1997). The Moroccan article reported that all the kappa values were very good (Kadri et al. 2005).

The American study had a delay of 1-2 days between the interviews (Sheehan et al. 1997). The Italian study had delay of 15 days between the first and second interview (Barbui C. et al. 2004). Although our mean delay was 1.8 days, the maximum delay between interviews was 6 days. A long delay between the first and second interview may lower the reliability. This is due to potential relief of symptoms after a few days in an acute psychiatric

ward, and thus fewer diagnoses are endorsed in the second interview. However, the first and second interview in the present study generated the same number of diagnoses.

High degree of psychopathology may also lower the reliability. With the exceptions of the psychosis and GAD sections, MINI is solely dependent on the patients' ability to report symptoms themselves. In an acute psychiatric ward patients are admitted due to acute crises. They may have difficulties in concentrating, thus giving inadequate answers. Some patients do not have insight in their psychopathology and may give wrong answers according to their actual psychopathology. Hence, diagnoses may be set more accurately on basis of unstructured clinical interviews and clinical intuition. In for example an outpatient clinic – or an acute psychiatric ward covering other populations - patients will have a better overview of their psychopathology and in general be calmer. This will probably make it easier to answer questions adequately and consistently and thus make the reliability and validity of MINI higher. Moreover, the psychopathology in an acute psychiatric ward may be just a “snapshot” of the crisis and thus give less information about the patients' diagnoses. However, the test-retest reliability in the present study is quite high. This may imply that the MINI diagnoses represent more than just a “snapshot”.

The French, Moroccan and Italian study included 42, 50 and 46 subjects respectively and can thus be compared to our study (Barbui C. et al. 2004;Kadri et al. 2005;Lecrubier et al. 1997). The American and Japanese articles included a higher number of subject (84 and 77 subjects respectively) (Otsubo et al. 2005;Sheehan et al. 1997) (The 42 subjects in the French study were included in the American study). A high base rate of diagnoses may make reliability higher. We have a relatively high level of co-morbidity in our population, making the amount of each diagnosis greater in relation to included subjects. Yet, the inclusion of more subjects and diagnoses in our study would probably have given a generally higher reliability.

Kappa values for OCD and GAD were low. The stem question of OCD is the most complex of all the questions in MINI. It contains long sentences that are hard to comprehend and easy to misinterpret and to mix with other symptoms, for example of psychosis. High degree of psychopathology probably makes it even harder to understand this question. Further on, this may produce answers that are incoherent in two different interviews and a low kappa value. Methodological differences may produce lower reliability. Differences in asking the patient to give an example, especially when they are not sure what to answer, can lower the diagnostic concordance. Asking the patient for examples when setting the OCD diagnose can probably in some cases be crucial. Perhaps we in a greater extent should have found a

common strategy. Or maybe the low kappa values for OCD actually say more about the way the questions are formulated; that the Norwegian version of these questions really is not reliable and should be reformulated. Again, this may just be the case when using MINI in an acute psychiatric ward.

The GAD MINI section is dependent on clinical judgement. Diagnose will not be set if the interviewer thinks that other diagnoses can explain the symptoms. This may represent a source of error and may explain the low kappa value. Divergence in clinical experience is a factor that may lower the reliability in general, and especially in this section. JM, which was the most experienced interviewer, endorsed significantly more GAD diagnoses than ØG.

Two of the sections in MINI partially depend upon clinical judgement. In the present study it seems that it works better in the psychosis section than in the GAD section. However, should it be inherent in other sections? We experienced that some of the obviously manic patients answered “no” to all of the questions in the current manic episode section, and kappa value for current manic episode was below 0.5. Perhaps clinical judgement should be integrated in this section. From a different perspective, one of MINI’s strengths is the non-academic profile. Its ability to give standardized diagnoses for research work would also have been weakened if more clinical judgement was implemented.

When assessing differences between ØG and JM concerning feasibility, patient acceptability and endorsement of diagnoses, most of the differences were non significant. Some of this lack of significance may represent type 2 bias due to the under powered profile of the present study.

This study explored test-retest reliability of the Norwegian MINI version. It would also be of great value to investigate the inter-rater reliability and the validity. It is appropriate to investigate and test MINI in other populations to explore rate of co-morbidity, feasibility, patient acceptance and reliability. This would also allow us to be more conclusive about the degree of psychopathology in Lovisenberg, and perhaps tell us if the weaknesses in the study are caused by methodological weaknesses or difficulties of interviewing an actual sicker population. It would be of interest to increase the number of subjects in this or further studies. Hence, the results will be more trustworthy and kappa values maybe higher.

Mini-International Neuropsychiatric Interview is the most used structured psychiatric interview in the world (cited Sep. 20, 2007; available from: www.medical-outcomes.com/HTMLFiles/MINI/MINI.htm). The present study indicated that the feasibility of the Norwegian MINI in an acute psychiatric ward was good enough to be satisfactory and not an argument against usage in this population. The kappa values were very good or good

(11 of 16) for most of the diagnoses indicating high test-retest reliability. The patient acceptance seemed very good. On the other hand, our reliability and feasibility results may suggest adjustments to make MINI fit better in an acute psychiatric ward. Maybe some of the sections or at least specific questions and considerations don't fit in this setting. The above discussion also implies the importance of methodological stringent thinking when testing the test-retest reliability. We could probably in a greater extent prior to the study have discussed the way to approach MINI's ambiguities. However, this should not overshadow that the results overall were very positive.

References

- Altman D.G 1991, "Some common problems in medical research," in *Practical Statistics for Medical Research*, First edn, Chapman and Hall, pp. 396-439.
- Anderson, D. M., Jefferson, K., Novak, P. D., & Elliott, M. A. 1994, *Dorland's illustrated medical dictionary*, 28th edn, W.B. Saunders Company.
- Barbui C., Ciuna, A., Nose' M., Levi, D., Patten, S. B., Amaddeo F., & Tansella, M. 2004, "The Reliability of the Mini-International Neuropsychiatric Interview - Italian Version", *Journal of Clinical Psychiatry*, vol. Volume 24, pp. 561-563.
- Blacker, D. 2005, "Psychiatric Rating Scales," in *Kaplan & Sadock's Comprehensive textbook of psychiatry*, Eighth edn, B. J. Sadock & V. A. Sadock, eds., Lippincott Williams & Wilkins, pp. 929-955.
- Friis, S. & Vaglum, P. 1999a, "Er målingene til å stole på?," in *Fra ide til prosjekt - en innføring i klinisk forskning*, Second edn, Tano Aschehoug, pp. 119-129.
- Friis, S. & Vaglum, P. 1999b, "Litt elementær statistikk," in *Fra ide til prosjekt - en innføring i klinisk forskning*, Second edn, Tano Aschehoug, pp. 111-118.
- Kadri, N., Agoub, M., El Gnaoui, S., Alami, K. M., Hergueta, T., & Moussaoui, D. 2005, "Moroccan colloquial Arabic version of the Mini International Neuropsychiatric Interview (MINI): qualitative and quantitative validation", *European Psychiatry*, vol. 20, no. 2, pp. 193-195.
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. C. 1997, "The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI", *European Psychiatry*, vol. 12, no. 5, pp. 224-231.
- Malt, U. F., Retterstøl, N., & Dahl, A. A. 2003a, "Den psykiatriske undersøkelsen," in *Lærebok i psykiatri*, Second edn, Gyldendal Norsk Forlag AS, pp. 117-136.
- Malt, U. F., Retterstøl, N., & Dahl, A. A. 2003b, "Psykiatriske forståelsesmodeller," in *Lærebok i psykiatri*, Second edn, Gyldendal Norsk Forlag AS, pp. 61-94.

Malt, U. F., Retterstøl, N., & Dahl, A. A. 2003c, "Psykiske lidelser - inndeling, utbredelse og behandlingsorganisasjon," in *Lærebok i psykiatri*, Second edn, Gyldendal Norsk Forlag AS, pp. 9-39.

Malt, U. F., Retterstøl, N., & Dahl, A. A. 2003d, "Stemningslidelser (Affektive lidelser)," in *Lærebok i psykiatri*, Second edn, Gyldendal Norsk Forlag AS, pp. 283-338.

Otsubo, T., Tanaka, K., Koda, R., Shinoda, J., Sano, N., Tanaka, S., Aoyama, H., Mimura, M., & Kamijima, K. 2005, "Reliability and validity of Japanese version of the Mini-International Neuropsychiatric Interview", *Psychiatry Clin.Neurosci.*, vol. 59, no. 5, pp. 517-526.

Pinninti, N. R., Madison, H., Musser, E., & Rissmiller, D. 2003, "MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance", *Eur.Psychiatry*, vol. 18, no. 7, pp. 361-364.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. 1998b, "The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10", *J Clin.Psychiatry*, vol. 59 Suppl 20, pp. 22-33.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. 1998a, "The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10", *J Clin.Psychiatry*, vol. 59 Suppl 20, pp. 22-33.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M. F., & Dunbar, G. C. 1997, "The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability", *European Psychiatry*, vol. 12, no. 5, pp. 232-241.

Thorsnæs, G., Løset, Å., Gundersen, N. H., Omre S.E., Skogstad, L., & Sundell, T. 2006a, "Helseforhold," in *Statistisk årbok for 2006*, Oslo kommune Utviklings- og kompetanseetaten, pp. 133-147.

Thorsnæs, G., Løset, Å., Gundersen, N. H., Omre S.E., Skogstad, L., & Sundell, T. 2006b, *Statistisk årbok for Oslo 2006* Oslo kommune Utviklings- og kompetanseetaten.

Zimmerman, M., Spitzer, R. L., Mezzich, J. E., & Berganza, C. E. 2005, "Classification in Psychiatry," in *Kaplan & Sadock's Comprehensive textbook of psychiatry*, Eighth edn, B. J. Sadock & V. A. Sadock, eds., Lippincott Williams & Wilkins, pp. 1003-1052.

Table 1: Cut-off values for interpretation of Kappa.

Value of κ	Strength of agreement
< 0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

Table 2: Number of diagnoses per interviewer (n=38)^a

Disorders	Jon (n=38)	Øystein (n=38)	P-value	Interview 1 (n=38)	Interview 2 (n=38)	P-value
Major depressive episode, current	25	26	NS	27	24	NS
Major depressive episode, recurrent	16	17	NS	16	16	NS
Dysthymia	2	3	NS	2	3	NS
Manic episode, current	3	8	NS	5	6	NS
Manic episode, past	7	7	NS	8	6	NS
Hypomanic episode, current	2	1	NS	2	1	NS
Hypomanic episode, past	6	0	0.011	2	4	NS
Panic disorder, current	12	11	NS	11	12	NS
Panic disorder, lifetime	19	18	NS	18	19	NS
Agoraphobia	13	17	NS	14	16	NS
Social anxiety disorder	11	13	NS	13	11	NS
Obsessive-compulsive disorder	7	6	NS	8	5	NS
Psychotic disorder, current	13	14	NS	13	14	NS
Psychotic disorder, lifetime	18	16	NS	17	17	NS
Mood disorder with psychotic features	2	4	NS	4	2	NS
Alcohol dependence	7	7	NS	7	7	NS
Alcohol abuse	2	0	NS	1	1	NS
Drug dependence	10	10	NS	11	9	NS
Drug abuse	1	0	NS	1	0	NS
Bulimia nervosa	1	2	NS	2	1	NS
Generalized anxiety disorder	12	3	0.009	6	9	NS
Total	189	183	NS	188	183	NS
Diagnoses per interview (Total/n)	5.0	4.8	NS	5.0	4.8	NS

^aAnorexia nervosa and anorexia nervosa, binge eating/purging type were excluded from analysis because none of the patients met MINI criteria for these 2 diagnoses. In assessing p-values NS means non significant difference.

Table 3: Test-Retest reliability of MINI (n=38)^a

Disorders	Kappa (SE)	Interview 1		
		Interview 2	++ -+	+ - --
Major depressive episode, current	0.82 (0.10)	24 0	3 11	
Major depressive episode, recurrent	0.68 (0.12)	13 3	3 19	
Dysthymia	0.36 (0.29)	1 2	1 34	
Manic episode, current	0.47 (0.20)	3 3	2 30	
Manic episode, past	0.65 (0.16)	5 1	3 29	
Panic disorder, current	0.81 (0.10)	10 2	1 25	
Panic disorder, lifetime	0.84 (0.09)	17 2	1 18	
Agoraphobia	0.67 (0.12)	12 4	2 20	
Social anxiety disorder	0.51 (0.15)	8 3	5 22	
Obsessive-compulsive disorder	0.17 (0.18)	2 3	6 27	
Psychotic disorder, lifetime	0.78 (0.10)	15 2	2 18	
Psychotic disorder, current	0.71 (0.12)	11 3	2 21	
Alcohol dependence	1.00 (0.00)	7 0	0 30	
Drug dependence	0.86 (0.09)	9 0	2 26	
Bulimia nervosa	0.65 (0.32)	1 0	1 35	
Generalized anxiety disorder	0.25 (1.83)	3 6	3 24	

^aDrug abuse, anorexia nervosa and anorxia nervosa, binge eating/purging type were excluded from analysis because none of the patients met MINI criteria for these 3 diagnoses. We did not have sufficient sample size to analyze hypomanic episode, current and past, alcohol abuse and mood disorder with psychotic features and these 4 diagnoses were excluded from analysis (None of the patients met MINI criteria in both interviews and we got a negative kappa value). Bulimia nervosa has a prevalence < 5 % and should be considered with care.

Table 4: Patient comments

Can I be interviewed again?

This should have been the first thing done during admittance. I missed depth on certain issues, not just yes/no.

Should have been more questions about physical pain. A bit subtle.

Should open up to more examples and ideas/visions (psychosis). Physical, smell/taste hallucinations. Get tired of it. Long enough.

Should be used as a routine interview by admittance. Performed by independent staff.

It may be difficult to give yes/no answers on most of the questions.

I felt that somebody understood my thoughts and worries.

Well thought through questions. Good questions = good answers.

Helpful in a way. Tried to be understood. A bit uncomfortable and tiring for the concentration.

If yes/no, short questions, concise questions, good if ill patients. Something in between.

One can always better the quality on this type of caring; the health services within the psychiatry!

Very interesting. Gave me a lot

Too little about interests etc.

Good lenght.

Perfect!

Very ok questions.

I would like this to be sent to my general practitioner.

Figure 1: Time to perform interview related to interview number. Interview number 1 is the start of the project and interview number 38 is the end of the project. Regression line for JM indicates that he had a more pronounced reduction in interview time than ØG from the start to the end of the project.

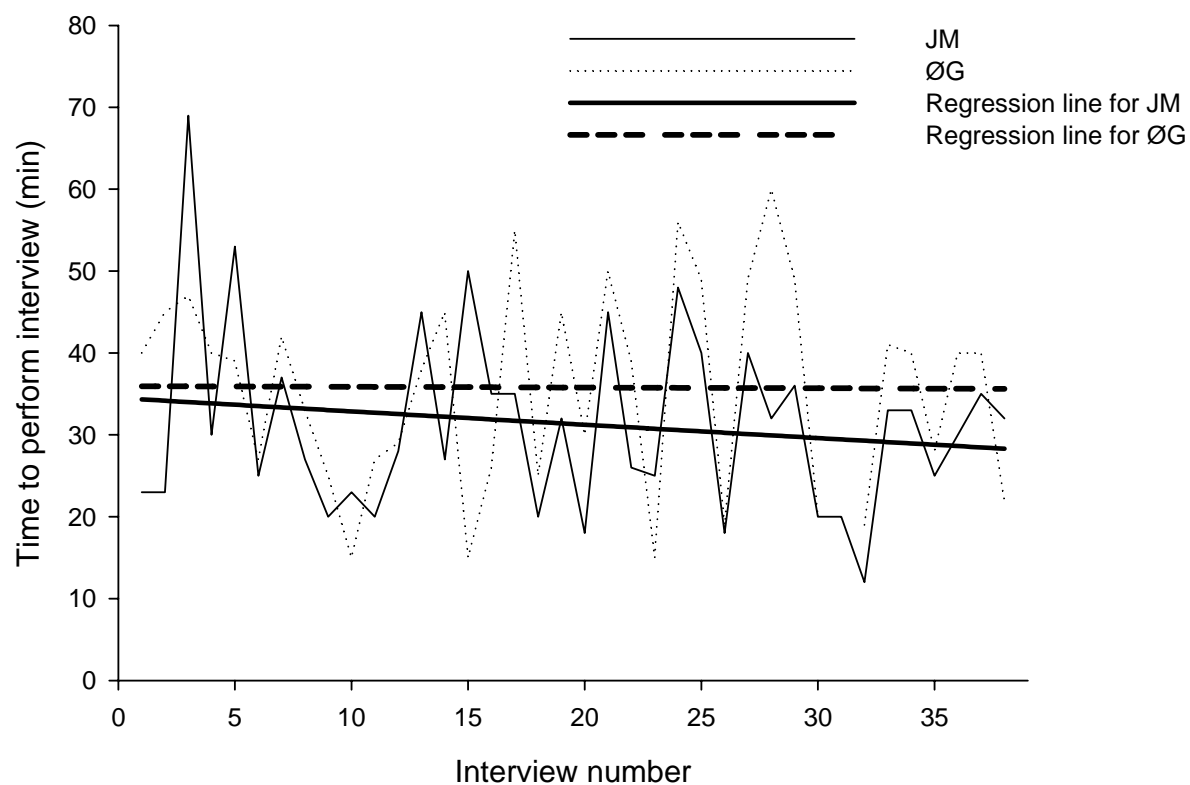
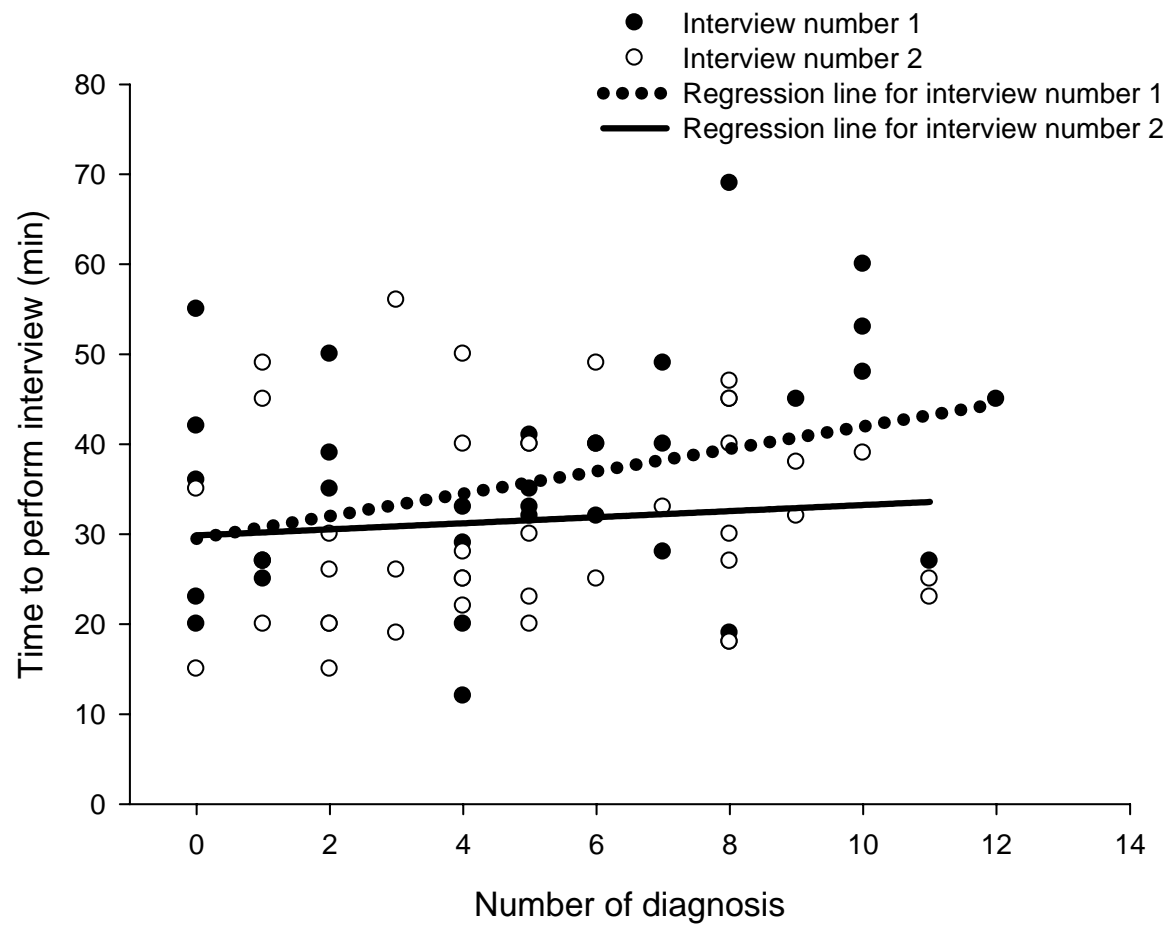


Figure 2

Time to perform interview related to number of diagnosis in each interview. Regression line for interview number 1 indicates that time to perform an interview increases along with number of diagnoses.



ACKNOWLEDGEMENTS

I would like to thank Jon Mordal and Jørgen Bramness for helping me with this project. They have been very supporting and have always been available for questions and guidance. I am very grateful to have worked with such enthusiastic and inspiring tutors.

I am thankful for the kind support from the staff at Lovisenberg Hospital, making the interviews possible. Thanks to Kjell-Petter Bøgwald for help with the patient acceptability forms and VAS-scales.

Øystein

Appendix 1:

Spørsmål til pasienten etter intervjuet

Vær vennlig å krysse av:

- 1) Hvor godt synes du spørsmålene i intervjuet dekket de plagene du for tiden sliter med?

|-----|

Svært dårlig/
Helt unyttig

Svært godt/
Veldig nyttig

- 2) Hvordan opplevde du å bli intervjuet på denne måten?

|-----|

Meget
ubehagelig

Meningsfullt/
behagelig

- 3) Andre kommentarer?

.....

**Takk for at du lot deg intervjuet i dette forskningsprosjektet
ved Lovisenberg Diakonale Sykehus!**

Vil du delta i forskningsprosjektet

"MINI i akuttpsykiatrien"?

Kjære pasient!

"MINI" står for Mini Internasjonalt Nervropsykiatrisk Intervju. Det er et strukturert intervju om psykiske plager som tar omtrent 30 minutter. Intervjuet er internasjonalt anerkjent og brukes også i norsk psykiatri, men det trengs mer kunnskap om hvordan MINI fungerer i Norge. Derfor spør vi om du vil bli med i et forskningsprosjekt som skal undersøke dette nærmere. Vi ønsker å intervju innlagte pasienter med MINI, for å kunne sammenlikne dette med tradisjonelle måter å kartlegge psykisk helse på. Noen pasienter vil bli intervjuet to ganger. Ved din deltakelse vil du gi et verdifullt bidrag til et prosjekt som ønsker å bidra til bedre utredning og behandling i akuttpsykiatrien.

Studieinnhold

- 1) Vi ber om tillatelse til å intervju deg om din psykiske helse med intervjuet "MINI". Intervjuene varer cirka 30 minutter og gjøres i løpet av oppholdet her.
- 2) Vi ber om tillatelse til å hente opplysninger fra din journal til bruk i forskningsprosjektet.

Datasikkerhet

- 1) Informasjon fra intervjuene med deg vil kun brukes til forskning, og blir ikke tilgjengelig for din behandler. Du har rett til å få innsyn i disse opplysningene og til å få noe endret hvis det er feil.
- 2) Medarbeiderne i prosjektet har taushetsplikt, og all informasjon om deg vil bli behandlet konfidensielt. Personlige opplysninger vil kun brukes til forskning og vil ikke kunne kobles til deg.

Risiko og nytte

Det er ingen risiko eller ubehag ved å delta i prosjektet. Mange pasienter vil oppleve utredningen som nyttig.

Frivillighet

Deltakelsen er frivillig og du behøver ikke bestemme deg med det samme. Du kan trekke deg fra prosjektet når som helst uten å oppgi grunn og uten at det får noen følger for behandlingen din. Da vil opplysningene om deg vil bli slettet, så lenge de ikke allerede er inngått i vitenskapelige arbeider.

Prosjektslutt

Noen pasienter vil bli kontaktet om et halvt år for en etterundersøkelse. Prosjektet avsluttes 31.12.2009. Da vil alle sensitive persondata bli slettet. Resultatene av studien vil bli publisert som gruppedata, uten at den enkelte kan gjenkjennes.

Prosjektledelse

Prosjektet er et samarbeid mellom Lovisenberg Diakonale Sykehus og Universitetet i Oslo. Det har ingen kommersielle formål. Forskere er medisinerstudent Øystein Gundersen og lege Jon Mordal. Avdelingen kan formidle kontakt med dem når som helst ved spørsmål eller uklarheter.

Bjørn Holm

Sjefslege

Lovisenberg D. Sykehus

Vibeke Lie

Avd. overlege, psyk.

Lovisenberg D. Sykehus

Lars Weisæth

Professor, undervisningsleder

Universitet i Oslo

Prosjektet er tilrådd av Regional komité for medisinsk forskningsetikk og av Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Samtykkeerklæring – prosjektet "MINI i akuttpsykiatrien":

Jeg har mottatt skriftlig og muntlig informasjon om prosjektet og sier meg villig til å delta.

Oslo, dato: _____

Signatur: _____

Plassér
pasientetiketten
her